Stat & Data Analysis 2.2 notes

CENTER & SPREAD

CENTER:

1. MEDIAN

Example: 3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20 Find the Median

QUARTILES:

Example: 3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20 Find Q1 & Q3

SPREAD:

RANGE:

IQR:

Same example: 3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20

Q3 = 5 IQR = Range =

Q1 = 12

**THE 5# SUMMARY:** Center & Spread together

**Visually: BOXPLOT**

Use the data from before: 3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20

Min = 3 Q1 = 5 Median = 7.5 Q3 = 12 Max = 20

Example: Create a boxplot of the following: heights of a sample of HS students

Min = 60" Q1 = 63 Med = 65.5" Q3 = 69 Max = 74"

**5 # Summary and Boxplots on the calculator:**

Find the 5# Summary & create a boxplot of the following data:

12 18 25 20 21 17 30 29 32 35

10 40 38 34 33 31 30 25 24 23

19 22 13 16 26 29 27 35 62 57

***\*\*See page 71-- 72 in the book***

***Statistics: STAT 🡪 CALC 🡪 1-var stats 🡪 ENTER 🡪 Find the list you want 🡪 ENTER***

**PRACTICE:**

1) Create a boxplot and find the 5# summary of the list TEST

2) Create a boxplot and find the 5 # Summary of the list GPA

***Finding outliers- the 1.5 x IQR test***

In order to formally to determine if there are outliers:

Example: Using the data from before, test for outliers

12 18 25 20 21 17 30 29 32 35

10 40 38 34 33 31 30 25 24 23

19 22 13 16 26 29 27 35 62 57

***5 # Summary:***

min = 10 Q1 = 20 Med = 26.5 Q3 = 33 Max = 62

***Another example:*** Use the list INCOM:

a) Find the 5# summary- write it down

b) Use this to do the outlier test. Find your acceptable range of data

c) Sort the list

(*STAT --> SortA --> ENTER --> find list INCOM --> ENTER)*

c) Put the list into your EDITOR WINDOW in your calculator

(*STAT --> EDIT --> find blank spot --> find list INCOM --> ENTER)*

d) Look at the highest and lowest points in the list. Then look at your outlier test- are there any outliers? If so, what are they?

***Another example:*** Use the list PAGAS:

a) Find the 5# summary- write it down

b) Use this to do the outlier test. Find your acceptable range of data

c) Sort the list

(*STAT --> SortA --> ENTER --> find list PAGAS --> ENTER)*

c) Put the list into your EDITOR WINDOW in your calculator

(*STAT --> EDIT --> find blank spot --> find list PAGAS --> ENTER)*

d) Look at the highest and lowest points in the list. Then look at your outlier test- are there any outliers? If so, what are they?

***Modified Boxplot:***

- If you decide that there are outliers, you can modify your boxplot

- Put a dot where the outliers are, and then make the whisker go to the next available point

- Try this with the example data from before:

12 18 25 20 21 17 30 29 32 35

10 40 38 34 33 31 30 25 24 23

19 22 13 16 26 29 27 35 62 57

Normal range: (0.5, 52.5)

**Example**: Create a boxplot (testing for outliers) of the following data. Put the data in L2 in your calculator.

10 45 60 62 55 50 49 51 52 55 69 67 63

64 66 65 70 72 73 48 46 47 53 54 57 58

59 60 70 8 60 56 47

**Parallel Boxplots:**

Example:

Example: Create parallel boxplots of the lists SATMF and SATMM (lists of SAT math scores of Males & Females).

**REVIEW:**

Center: Spread:

Calculator: Picture:

***The other center…***

CENTER:

1. MEAN

SPREAD:

STANDARD DEVIATION

On the calculator:

Example: Look at the two sets of data and decide which one has the larger std. deviation:

1) {4, 9, 12, 17, 20} or {4, 6, 7, 8, 10}

2) {100, 140, 150, 160, 200} or {10, 50, 60, 70, 110}

Examples: Which one has the larger std. dev?

1) 4, 5, 6, 6, 7, 7, 7, 8, 8, 9, 10

2) 4, 4, 4, 5, 5, 6, 7, 8, 9, 9, 10, 10, 10

3) 1, 3, 5, 7, 7, 9, 9, 9, 11, 11, 13, 15, 17

4) 5, 5, 5, 5, 5

**Properties of the Standard Deviation:**

SUMMARY:

Mean vs. Median (in distributions)

\* Symmetric:

mean = median

\* Left skewed:

mean < median

\* Right skewed:

mean > median

Examples:

1) Mean = 25

Median = 50

2) Mean = 300

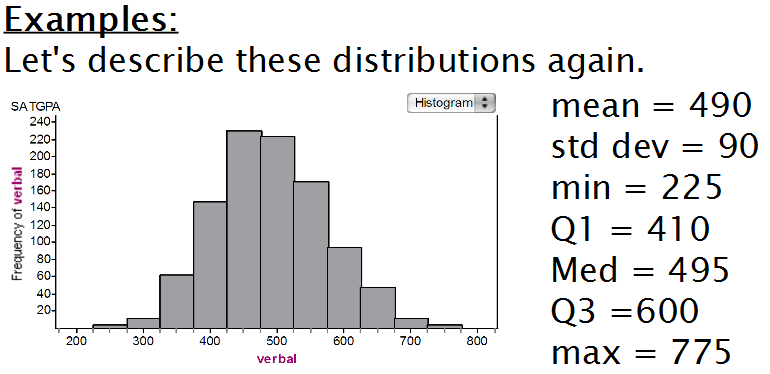
Median = 305

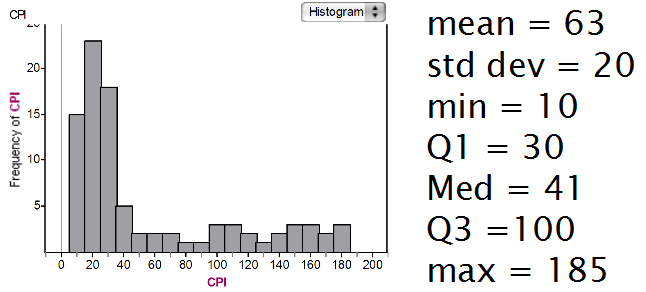
3) Mean = 240

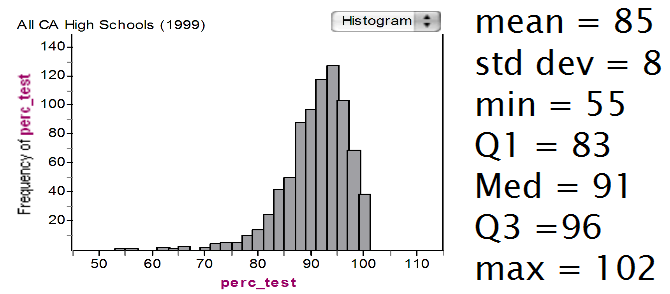
Median = 200

* The mean is \_\_\_\_\_\_\_\_\_\_\_\_\_ by outliers
* The median is \_\_\_\_\_\_\_\_\_\_\_\_\_\_ to outliers

So which do we use, and when?







**Comparing Distributions:**

* When comparing 2 distributions, you must still…
* You must also …
* Use

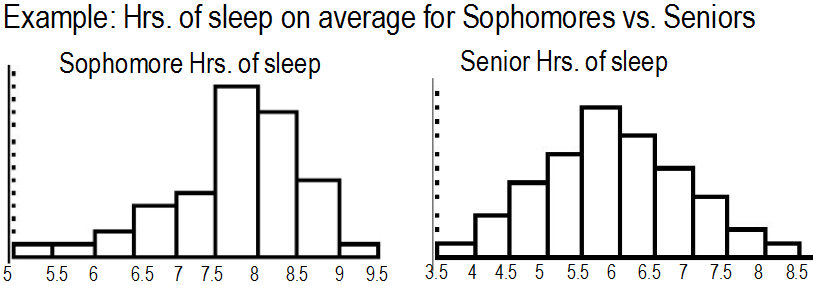
*Example: Block 2 & 3 took a test (out of 100 points).*

\* The shape of Block 2 is \_\_\_\_\_ which is similar to/different than the shape of Block 3 which is \_\_\_\_\_\_. Both are unimodal.

\* The center of Block 2 is the mean/median of \_\_\_\_ which is higher/lower than the mean/median of Block 3 which is \_\_\_\_.

\* The IQR/std. dev. of Block 2 is \_\_\_\_\_\_ which is wider/smaller than the IQR/std. dev. of Block 3 which is \_\_\_\_\_\_\_. The range of Block 2 is (\_\_, \_\_) which is wider/smaller than the range of Block 3 which is (\_\_\_, \_\_\_).

\* Other??? outliers? gaps?



*M = 7.3 M = 6*

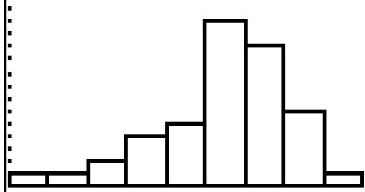
*IQR = 3 IQR = 3.5*

*mean =6.6 mean = 6.2*

*s = 0.8 s = 0.9*

Example: Compare and describe the SATMF & SATMM lists. Create the histograms (on your calculator only) & find the 1-Var Stats before you start.

Example: Compare and describe the BLK1 & BLK2 lists. These are quiz scores out of 20 points. Create the histograms (on your calculator only) & find the 1-Var Stats before you start.



**OUTLIERS: How do they affect summary statistics?**

If I add an **upper** outlier, how will it affect the summary stats?

Mean Median range IQR std. dev,

If I add a **lower** outlier, how will it affect the summary stats?

Mean Median range IQR std. dev,

**\* Outliers do not really affect the Median, IQR, or Quartiles**

**\* Outliers STRONGLY affect the mean, standard deviation, and range**

**Example:** Workers at a local business make on average $30,000 per year. A new employee is hired to be the company President, and he will make $70,000 per year. What will happen to the following summary statistics if the President's salary is added to the group?

Mean Median range IQR std. dev,

**Example:** Workers at a local business make on average $30,000 per year. A new employee is hired to be part time only and he will make $10,000 per year. What will happen to the following summary statistics if the new employee's salary is added to the group?

Mean Median range IQR std. dev,